

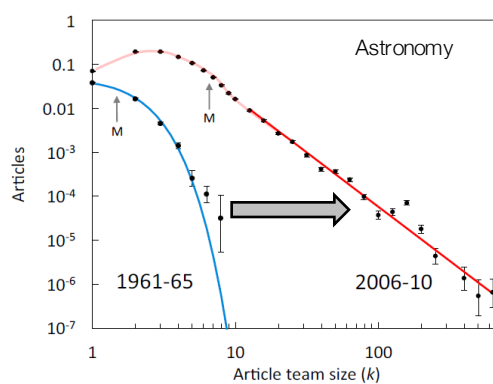
The role of small teams in big science era

Staša Milojević

School of Informatics and Computing, Indiana University Bloomington

Modern science: the rise of collaboration

- Collaboration/team work is one of the defining features of modern science
- The increasing level of knowledge production by large science teams



1960s:
Individual authors (90%) and
small teams

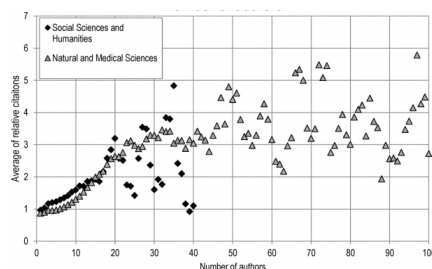
2000s:
Individual authors (<10%)
Mean team size ~7
“Big Science” papers on the
rise

Previously only high-energy
physics and biomedical; now
most experimental fields have
a “Big Science” component

Milojević, S. (2014). *PNAS*, 111(11), 3984

The role of small teams in big science era

- Big science results are more visible (Human Genome Project, CERN, ...)
- Some efforts require very large teams
- On average, papers with more authors are cited more
 - True even for few authors
 - No change beyond 30 authors
- Small teams <10
- Large teams 10-100
- Big science 100-1000



Larivière et al. (2014)



Is small-team mode of knowledge production a thing of past?





Is society wasting resources by continuing to fund individual researchers who primarily work alone or with only a few collaborators?



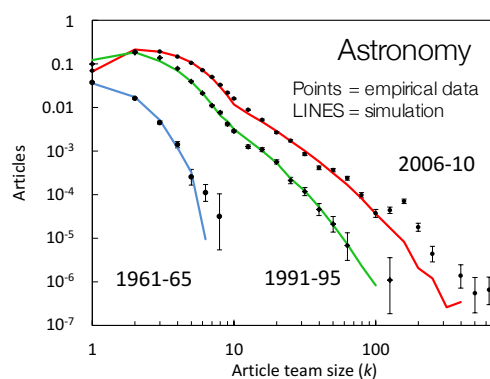
Model: how the teams of different sizes emerge and develop?

- New model for team sizes:
 - For each team there is a team **leader**
 - Teams are drawn from a Poisson distribution characterized by some mean
 - Teams formed in this way are **core teams**
 - Core teams applied to standard research questions
 - Core teams don't change in time except when someone retires [simplification of what actually happens]

Model: how the teams of different sizes emerge and develop?

- New model for team sizes:
 - In **parallel** with core team the lead author assembles **extended team**
 - The role of the extended team is to address certain type of research questions for which core teams are not sufficient
 - Initially the extended team is just the core team
 - In the model each time the lead author produces a paper there is a probability s/he will use core or extended team
 - On subsequent uses of the extended team it is allowed to grow in proportion to **the productivity of the existing extended team members**
 - More productive teams will grow faster (cumulative advantage)
 - Cumulative advantage can lead to a power-law distribution

Model reproduces the change in team size distribution

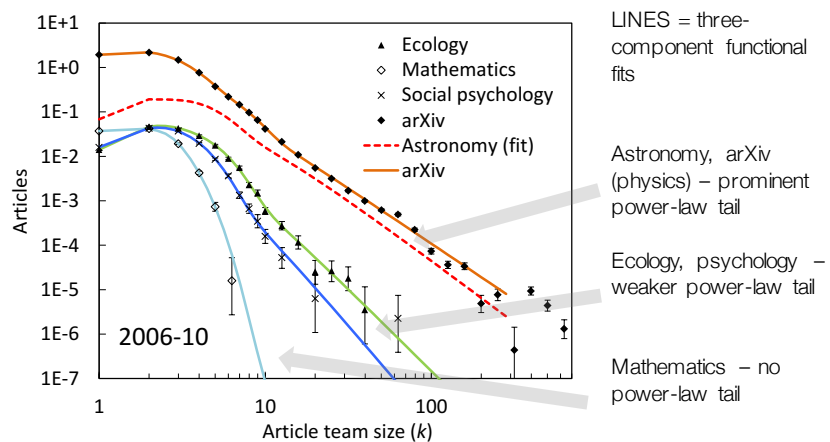


Every large team **originates** from a small team.

While many small teams stay small, some quickly accumulate additional members proportionally to the past **productivity** of team members, developing into larger teams, and allowing them to grow even faster.

Milojević, S. (2014). *PNAS* 111(11), 3984

Allows prediction of the change towards big science in other disciplines



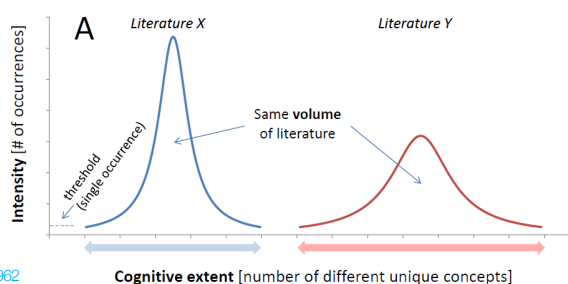
Miljević, S. (2014). *PNAS* 111(11), 3984

Is the only role of small teams to seed big ones?

- Which approach contributes more to modern science intellectually?
- Are the efforts of smaller teams becoming intellectually obsolete?
- Answer should not be based only on the number of papers that big teams produce, or citations, but on measures of **cognitive content and diversity** of work
- Currently no metric for cognitive content

Big-data method for measuring the cognitive extent of scientific literature

- Get concepts from parsing titles into phrases
- Idea: count the number of unique phrases in unit quotas of ~1000 articles
- 1000 articles which contain fewer unique phrases (X) have a smaller cognitive content than 1000 articles with more diverse phrases (Y)
- Large measurement quota is needed to reduce the stochasticity of the titles

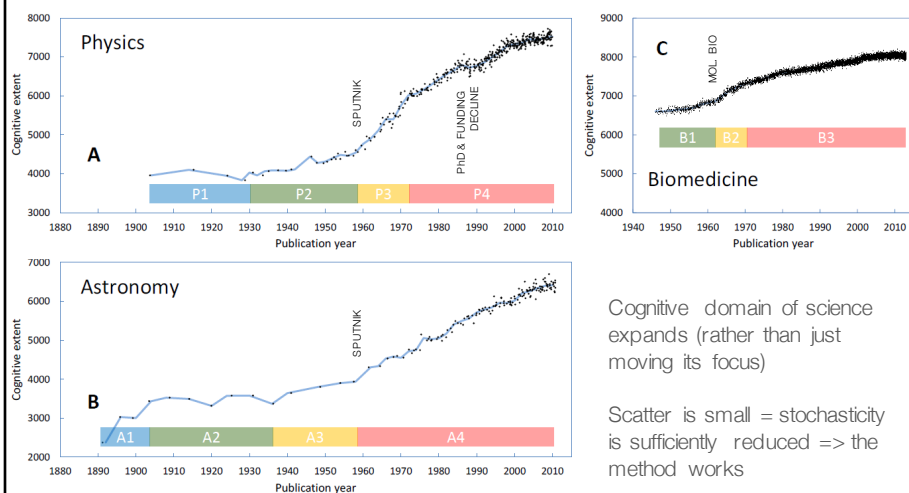


Milojević, S. (2015). *Journal of Informetrics* 9, 962

Cognitive concepts from titles

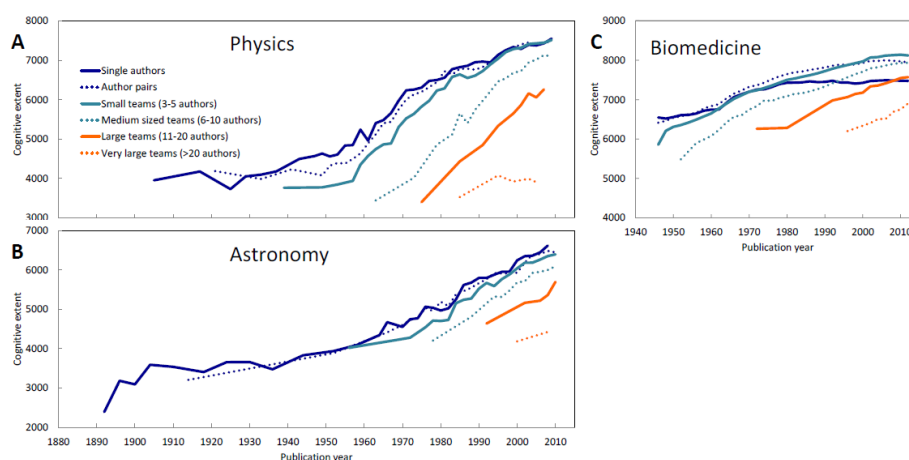
- Phrases identified automatically
- Use general words to separate phrases
- General words: prepositions, articles + non-specific words, such as:
study, analysis, result, determination, comparison, discovery
- Examples of titles with phrases capitalized:
 - **HALOS and VOIDS in MULTIFRACTAL+MODEL of COSMIC+STRUCTURE**
 - **COLLISION+STRENGTHS for ELECTRON+IMPACT+EXCITATION of FINE+STRUCTURE+LEVELS in FE+XIII**
- Old words in a new context:
 - **MASS, ENERGY -> MASS-ENERGY EQUIVALENCE**
 - Method identifies it as a new phrase (introduced in relativity theory)

Evolution of the cognitive extents of physics, astronomy, and biomedicine



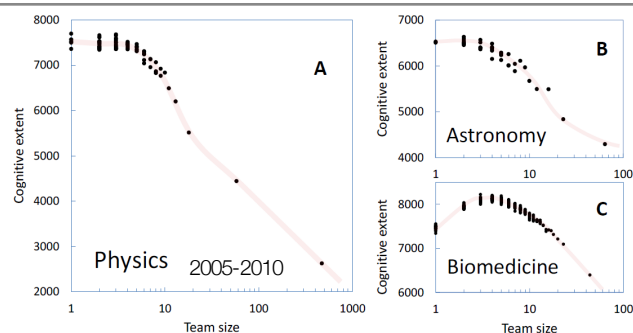
Milojević, S. (2015). *Journal of Informetrics* 9, 962

Evolution of the cognitive extents – by teams of different sizes



Milojević, S. (2015). *Journal of Informetrics* 9, 962

Inverse correlation between research team size and the cognitive extent of scientific output



Miljević, S. (2015). *Journal of Informetrics* 9, 962

In physics and astronomy, single authors, pairs of authors, and small teams cover the largest intellectual territory, the same size as the entire field. Larger teams cover significantly smaller cognitive territory. In biomedicine, the small teams (3-5 authors) cover the largest domain, but as in astronomy and physics, the very large teams cover the smallest cognitive territory.

Plus: Topics covered by large teams are not exclusive to them

Conclusions

- Even today, the small teams remain the necessary seeds for the formation of larger teams.
- The large teams are more specialized in nature and consequently do not encompass the intellectual breadth of the entire discipline covered by small teams.
- Thus the small teams appear to be critical in maintaining the intellectual diversity and expanding the frontiers of science, and may serve as the incubators for the topics that big teams work on.

Thank you!

Staša Milojević
School of Informatics and Computing
Indiana University Bloomington
smilojev@indiana.edu